Genetic evidence points to distinct paternal settlers of the Faroe Islands and Iceland

Allison E. Mann[1,2†], Eyðfinn Magnussen[3] and Christopher R. Tillquist[2*†]

[1]Department of Anthropology, University of Wyoming, Laramie, WY, United States, [2]Department of Anthropology, University of Louisville, Louisville, KY, United States, [3]Faculty of Science and Technology, University of the Faroe Islands, Tórshavn, Faroe Islands

Introduction: The Faroe Islands are a small archipelago located in the North Atlantic likely colonized by a small group of founders sometime between 50 and 300 CE. Post colonization, the Faroese people have been largely isolated from admixture with mainland and other island populations in the region. As such, the initial founder effect and subsequent genetic drift are likely major contributors to the modern genetic diversity found among the Faroese.

Methods: In this study, we assess the utility of Y-chromosomal microsatellites to detect founder effect in the Faroe Islands through the construction of haplotype networks and a novel empirical method, mutational distance from modal haplotype histograms (MDM), for the visualization and evaluation of population bottlenecks.

Results: We compared samples from the Faroe Islands and Iceland to possible regional source populations and documented a loss of diversity associated with founder events. Additionally, within-haplogroup diversity statistics reveal lower haplotype diversity and richness within both the Faroe Islands and Iceland, consistent with a small founder population colonizing both regions. However, in the within haplogroup networks, the Faroe Islands are found within the larger set of potential source populations while Iceland is consistently found on isolated branches. Moreover, comparisons of within-haplogroup MDM histograms document a clear founder signal in the Faroes and Iceland, but the strength of this signal is haplogroup-dependent which may be indicative of more recent admixture or other demographic processes.

Discussion: The results of the current study and lack of conformity between Icelandic and Faroese haplotypes implies that the two populations were founded by different paternal gene pools and there is no detectable post-founder admixture between the two groups.

FIGURE 1
Geographic location of the Faroe Islands and the three cities where the samples were collected.

1 Introduction

There was a man named Grim Camban. He first settled the Færeys in the days of Harold Fairhair. For before the king's overbearing many men fled in those days. Some settled in the Færeys and began to dwell there, and some sought to other waste lands. Aud the deeply wealthy fared to Iceland, and on her way thither she came to the Færeys, and there she gave Olof the daughter of Thorstan the Red in marriage: whence is come the greatest lineage of the Færey–folk, whom they call the Gate-beards, that dwell in Eastrey (Powell, 1896).

The Faroe Islands consist of an archipelago of 18 small islands, located in the North Atlantic, between South Norway, Iceland and Scotland (Figure 1). As a result of their demographic history and relative geographic isolation, the Faroe Islands, along with other North Atlantic Island populations, are genetically homogenous as compared to mainland populations (Wilson et al., 2001; Helgason et al., 2003; Goodacre et al., 2005). Historical and archaeological sources report that the Faroe Islands were settled around 800 CE by Vikings largely from western Norway (Arge, 1991; Arge et al., 2005). However, a growing body of evidence suggests that the islands were settled prior to this time, possibly by Celtic monks or other persons originating from the British Isles (Magnusson, 2003). Carbon dating of peat moss and barley grains support two pre-Viking settlement phases, around 300–500 CE and 500–700 CE (Church et al., 2013). More recently, Curtin et al. (2021) detected sheep-DNA in archaeological sediments from 500 CE, and based on modern whole-genome data, Gislason (2023) estimated that the initial founding of the Faroe Islands occurred between 50 and 300 CE, potentially two to three centuries earlier than previously believed based on archaeological findings alone.

TABLE 1 Probability that a mutation will occur in any given generation (μ) for each locus typed in this study.

Locus	μ (%)	References
DYS19	0.25	NIST 2012
DYS385a	0.21	NIST 2012
DYS385b	0.21	NIST 2012
DYS388	0.10	Ravid-Amir and Rosset, 2010
DYS389I	0.24	NIST 2012
DYS389b	0.35	NIST 2012
DYS390	0.25	NIST 2012
DYS391	0.28	NIST 2012
DYS392	0.07	NIST 2012
DYS393	0.08	NIST 2012
DYS426	0.07	Ravid-Amir and Rosset, 2010
DYS438	0.07	NIST 2012

Genetic evidence suggests that, like other human populations on North Atlantic islands (including Iceland), the maternal and paternal founding populations of the Faroe islands are asymmetrical, with the vast majority of Faroese mitochondrial genomes being Celtic in origin, while most Y-chromosomal haplogroups are Scandinavian in origin (Jorgensen et al., 2004; Als et al., 2006). In addition, enslaved individuals from the British Isles of both sexes may have accompanied the early Faroes settlers and thralled Irish, Slavic, and Scandinavian individuals played an important role in early Viking society (Johnston, 1975; Jones, 1984). While it is unknown how large the

founding population was, the demographic history of the Faroe Islands suggest that the islands were initially colonized by a small group of founders with a slow, but steady, increase in the population over time. During a 300-year period, from 1300 to 1600, the Faroese population was quite stable, around 3,000 to 4,000 people (Strøm, 2017). In 1600, it was estimated to be 3,200, but in 1650, it decreased to 2,515 individuals (Strøm, 2017). However, after 1700 there was a constant growth in the Faroese population; in 1700, around 4,000 people lived in the Faroes Islands, 5,255 in 1801, and 15,230 in 1901 (Strøm, 2017). In 2020, the population had reached 45,749 individuals. However, like other countries, the Faroe Islands have also been affected by different epidemics, which has caused the population to fall. During the years 1720–1865, the annual mortality rate (death per 1,000 inhabitants) fluctuated between seven and 42 (Strøm, 2017) (See Supplementary Figure S1 for more information).

To identify signatures of founder events in the Faroese population, we analyzed 12 individual Y-chromosomal microsatellite loci from 139 Faroese males (Table 1) and classified each haplotype into haplogroups using a Bayesian classification method (Athey, 2006). As Y-chromosomal haplogroups are defined by relatively slowly mutating single nucleotide polymorphisms (SNPs) and Y-chromosomal haplotypes are defined by relatively quickly mutating microsatellites, short tandem repeats (STRs) (Zhivotovsky et al., 2004; Xue et al., 2009), we expect that haplotype diversity will recover faster than haplogroup diversity after a founding event. The fact that the vast majority of the Y chromosome is not subject to recombination indicates that hierarchical analysis of quickly mutating loci within groups delineated by slowly mutating loci may elucidate certain observations (de Knijff, 2000). Specifically, by definition, the occurrence of any SNP on the Y chromosome generates a novel haplogroup, the STR haplotype of which is singular, effectively resetting STR variation to zero. If the haplogroup is maintained in a population and increases in frequency, in subsequent generations haplotype variability will accumulate; in the absence of excessively high STR mutation rates, the original haplotype can be reliably inferred from modal values (de Knijff, 2000).

This study undertakes to further document the genetic impact of the colonization of the Faroe Islands. By using an extended number of STR loci for increased resolution, and by analyzing haplotype diversity in the context of haplogroups, a more precise, nuanced, and accurate understanding emerges that clearly differentiates the evolutionary impact of the origin processes on the respective populations and permits exclusion of substantial post-founding paternal gene flow between Iceland and the Faroe Islands.

2 Materials and methods

2.1 Sample collection, DNA extraction, and sequencing

Buccal cells were collected using sterile swabs in 2004 and 2005 from 139 Faroes males living in three geographically disparate locations within the island group: (1) the city of

Klaksvík (n = 41) in the northern islands, (2) Tórshavn (n = 48) in the central islands, and (3) Tvøroyi (n = 50) in the south (Figure 1). DNA was isolated from swabbed buccal cell samples using a standard phenol-chloroform extraction protocol. Each DNA sample was next scored for 12 different Y-chromosomal STR markers (DYS19, DYS385a, DYS385b, DYS388, DYS389I, DYS389II, DYS390, DYS391, DYS392, DYS393, DYS426 and DYS438) in two multiplexes as described in Quintana-Murci et al. (2004) with primers and PCR conditions given in Redd et al. (2002). DYS389b was calculated by subtracting DYS389I from DYS389II (Kayser et al., 1997). Since DYS385 is characterized by a duplication, the shorter allele was assigned to DYS385a and the longer to DYS385b within each individual haplotype (Pacheco et al., 2005). To find potential parental sources to the Faroese population, buccal swabs from 47 Norwegian and 36 Swedish males were collected in 1999 for the current study and were prepared in an identical manner as described above. Additionally, previously published haplotype data from Iceland (n = 181) (Helgason et al., 2000), Denmark (n = 185) (Hallenberg et al., 2005), and Ireland (n = 148) (Ballard et al., 2006) were used as additional source populations (Supplementary Material S1). Individuals within the putative parental data sets that had non-discrete or missing values were not included for the purposes of compatibility with the Faroese samples.

2.2 Haplogroup inference

Haplogroups were inferred from haplotype data using an allele frequency-based Bayesian method (Athey, 2006), using nomenclature for haplogroups as defined by the Y-Chromosome Consortium (YCC) (Karafet et al., 2008). The online haplogroup prediction module (http://hprg.com/hapest5) allows the user to input specific marker values and returns a goodness of fit and probability score assigning an individual haplotype to one of ten common European haplogroups. In cases where individual haplotypes did not fall into expected Scandinavian or Celtic haplogroups (i.e., not R1a, R1b, or I1: n = 74 haplotypes), they were compared to the Y-Chromosome Haplotype Reference Database (YHRD) (https://yhrd.org) (Willuweit and Roewer, 2015).

Reliability of the Athey method of haplogroup inference was assessed by comparing Norwegian and Swedish samples that had been SNP-typed for I-P38 (I1*), I-P37b (I1b), I-P40 (I1a1), J-12f2a (J*), J-M172 (J2), N-tat (N3), R-M173 (R1*), R-P25 (R1b), R-SRY10831b (R1a) (The Y Chromosome Consortium, 2002). All Swedish samples were placed in the correct haplogroup. All Norwegian samples were also scored correctly save one SNP-tested R1a haplotype which scored a 62.3% probability for R1b and a 37.3% probability for R1a. Given its previous SNP testing, this sample was grouped with other R1a haplotypes for downstream analysis. Of all unique haplotypes included in this study (n = 535) only 5% had a probability score of less than 90% (n = 27). STR loci DYS388, DYS426, and DYS438 were only used for haplogroup inference, and are not included in subsequent analyses, as one or more were not reported in the complete Norwegian, Irish, Swedish, or Danish datasets.

2.3 Haplotype network inference

Network analysis using high-resolution allele frequency data can parse distinct haplotypes within the same haplogroup (Zerjal et al., 2003; Moore et al., 2006). For each major haplogroup (R1a, R1b, I1), we generated haplotype networks using pairwise distances between all individuals (Hamming distance) generated with the R libraries adegenet (v2.1.10) (Jombart, 2008) and poppr (v2.9.5) (Kamvar et al., 2014). Correspondence analysis of all haplotypes across all haplogroups was performed with the R libraries adegenet (v2.1.10) (Jombart, 2008) and vegan (v2.6–4) (Oksanen et al., 2019). We calculated the degree of genetic differentiation among and between populations (Phi-statistics) within our haplotype networks using poppramova and the randtest function in the ade4 (v1.7–22) R library (Dray and Dufour, 2007).

2.4 Histograms of within-haplogroup mutational distances

Assuming a stepwise mutation model (Kimura and Ohta, 1978), we estimated mutational distances from the modal haplotype (MDM) for each haplogroup. We define the MDM as the sum of the absolute number of repeat differences at each locus that an individual's haplotype deviates from the population-specific modal haplotype which itself is determined by the modal repeat count at each locus. The MDM for each individual is therefore calculated

as: $\sum_{i}^{l} | Mo_i - x_i|$

Where Mo_i is the mode for any given locus within a haplogroup, l is the total number of loci, and x is the haplotype of an individual at locus i. This approach is an empirical extension of the method described by Rogers and Harpending (1992) who used mitochondrial data to calculate mismatch distributions. Mismatch distributions of pairwise sequence differences have been shown to reproduce the structure of phylogenetic trees in simulated data (Slatkin and Hudson, 1991), and have been used to document expansions of human populations (Di Rienzo and Wilson, 1991). Specifically, the overall shape of the mismatch distribution and the position of the central tendency on the abscissa reflects time-since-expansion (Rogers and Harpending, 1992). This observation accords with the finding that in an exponentially growing population the expected phylogeny of haplotypes will be star-like and close to the coalescent (Slatkin and Hudson, 1991). We hypothesize that different distribution patterns around modal haplotypes indicate various demographic processes involving population size, age, and the relative success of individual lineages. In the case of recent founder effect, we expect that most mutational distances from the modal haplotype will be a single mutational step away or fewer which we have termed "neighbor haplotypes". Specifically, we expect our MDM histogram analysis to represent the accumulation of within-population microsatellite diversity on a given haplogroup background since the time of the founding event. In the context of the expansion of an island population, predominant haplotypes in the modern population likely represent early founders. Consider a case where the small number of earliest founding males are paternally related; all will share identical (or near identical) Y-chromosomal haplotypes. During subsequent demographic

expansion, the accumulation of mutations on founding lineages will be reflected in a progressive loss of zero mutation step instances in MDM histograms resulting in a gradual progression of the central tendency to the right.

In addition to the proportion of neighbor haplotypes, we calculated the degree of skewness (γ1) for each histogram using the R moments library (v 0.14.1) (Komsta and Novomestky, 2022) wherein a value of zero indicates complete normality, a positive value indicates a positive skew, and a negative value a negative skew. While empirical mismatch distributions do not mirror simulated populations at equilibrium, they reflect those of recently expanded non-equilibrium populations with reasonable accuracy (Rogers and Harpending, 1992). As our MDM histograms are generated using a modified pairwise comparison procedure, they should document similar distributional expectations with regards to time-since-founding and changes in demography. Given our interest in detecting a loss of diversity within haplogroups, comparing haplotypes to a locus-specific mode is a more appropriate methodology for our purposes.

All analysis scripts are available at https://github.com/aemann01/faroes_y. Scripts to calculate population specific modal haplotypes and generate MDM histograms are available as an interactive Shiny app at https://aemann01.shinyapps.io/mdmhistogram/.

2.5 Diversity estimates

Finally, we calculated metrics of genotypic diversity, richness, and haplotype diversity using the R poppr (v 2.9.5) (Kamvar et al., 2014) and pegas (v 1.3) (Paradis, 2010) packages. We reported four major diversity statistics across all haplogroups in all populations as well as haplogroup specific diversity metrics. Reported metrics include (1) MLG: the number of multilocus genotypes in each population, (2) eMLG: the expected number of MLG with a minimum of 10 samples after rarefaction, (3) lambda: a sample size corrected Simpson's index (Simpson, 1949) calculated as: $((N/(N-1))*lambda)$ where N equals the number of individuals in the population and lambda is the uncorrected Simpson's index value, and (4) haplotype diversity (Hd). Hd ranges from zero to one where zero indicates no haplotype diversity and one complete haplotype diversity (Paradis, 2010).

3 Results

3.1 Faroese Y-chromosomal haplogroups

The most common haplogroups in all samples investigated here are haplogroups R1a, R1b, and I1 (Figure 2A). Among the Faroese sample, these three haplogroups constitute 42% (n = 58), 25% (n = 35), and 21% (n = 29) of the dataset, respectively (Figure 2A). Rare haplogroups were also found at low frequency. For example, haplogroup J1 constituted 4% of all Faroese samples (n = 6), Q 3% (n = 4), and haplogroups E1b1b (n = 2), I2b1 (n = 1), I2b(xI2b1) (n = 1), I2a (xI2a1) (n = 1), L (n = 1), and N (n = 1) combined make up approximately 6% of all Faroese samples (Table 2). Rare haplogroups found in source populations include G2a (n = 5),
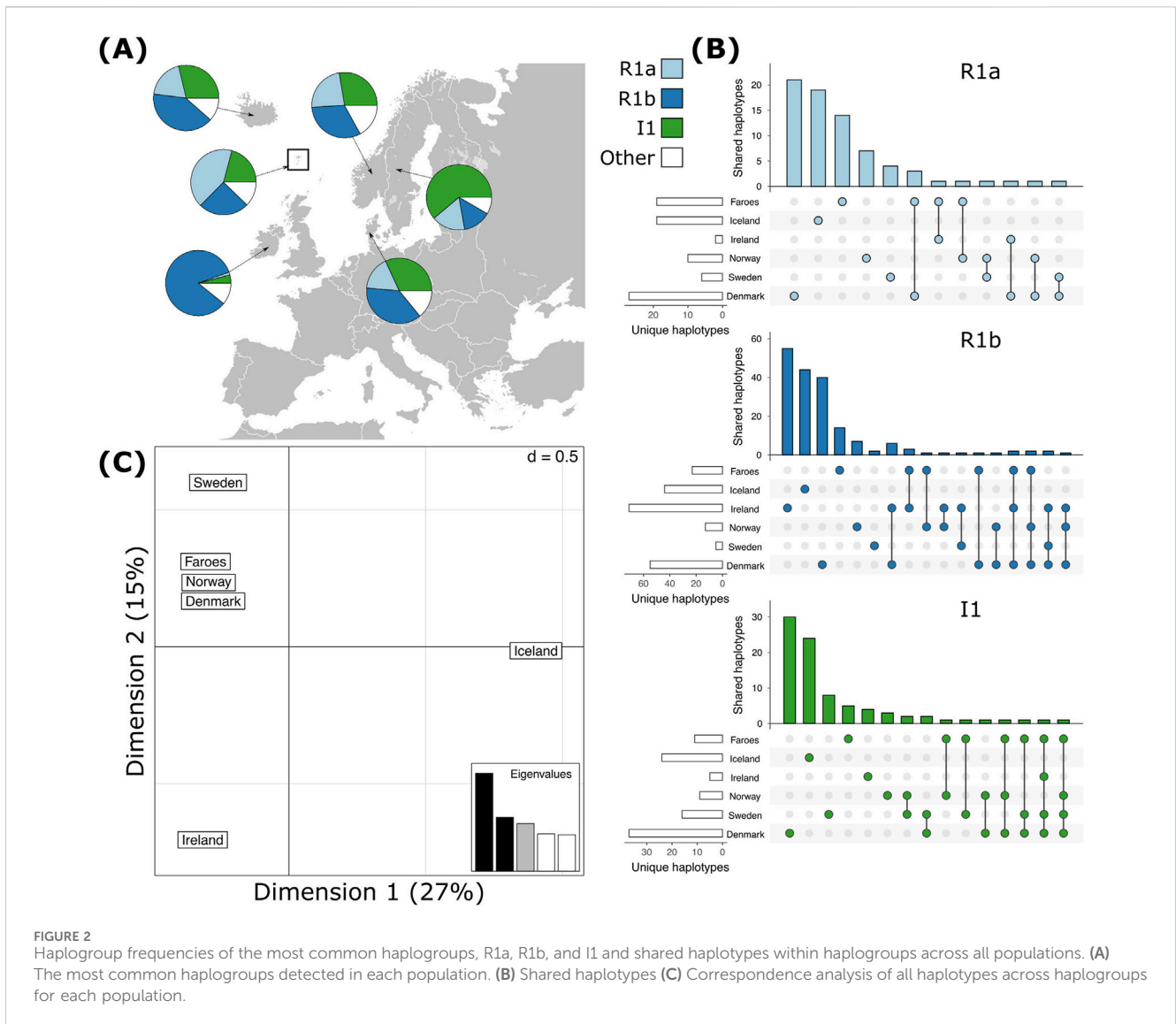
FIGURE 2
Haplogroup frequencies of the most common haplogroups, R1a, R1b, and I1 and shared haplotypes within haplogroups across all populations. **(A)** The most common haplogroups detected in each population. **(B)** Shared haplotypes **(C)** Correspondence analysis of all haplotypes across haplogroups for each population.

TABLE 2 Frequencies of different haplogroups in people living in Scandinavia and Ireland.

Population	R1a	R1b	I1	J1	Q	E1b1b	I2b1	I2b	I2a	L	N	Other[a]
Faroes	0.42	0.25	0.21	0.04	0.03	0.01	0.01	0.01	0.01	0.01	0.01	0.00
Klaksvík	0.37	0.27	0.15	0.10	0.02	0.05	0.00	0.02	0.02	0.00	0.00	0.00
Tórshavn	0.42	0.23	0.25	0.00	0.04	0.00	0.02	0.00	0.00	0.02	0.02	0.00
Tvøroyri	0.48	0.24	0.22	0.04	0.02	0.00	0.00	0.00	0.00	0.00	0.00	0.00
Denmark	0.17	0.37	0.32	0.00	0.02	0.02	0.05	0.01	0.01	0.00	0.01	0.03
Iceland	0.19	0.40	0.29	0.00	0.00	0.07	0.02	0.00	0.00	0.00	0.00	0.03
Ireland	0.01	0.84	0.04	0.00	0.00	0.01	0.05	0.00	0.03	0.00	0.00	0.02
Norway	0.23	0.32	0.28	0.00	0.02	0.00	0.06	0.00	0.00	0.00	0.06	0.02
Sweden	0.17	0.14	0.61	0.00	0.03	0.00	0.03	0.00	0.00	0.00	0.03	0.00

[a]Other haplogroups in Denmark include: G2a, J2a1b, J2b. Iceland: G2a, Ireland: G2a, J2b. Norway: J2a1h.

TABLE 3 Total Estimates of Diversity. Estimates of diversity across all haplotypes independent of haplogroup assignment for each population.

Population	N	MLG	eMLG	λ	Hd	k	%Unique	%Private
Faroes	139	64	25.20	0.97	0.97	68	48.9	30.2
Denmark	185	144	33.70	1.00	1.0	163	88.1	58.9
Ireland	148	92	29.70	0.98	0.98	100	67.6	49.3
Iceland	181	98	28.40	0.98	0.98	120	66.3	49.7
Norway	47	40	31.60	0.99	0.99	40	85.1	48.7
Sweden	36	30	30.00	0.99	0.99	30	83.3	47.2
Total	736	468	34.20	1.00	1.0	521		

N indicates the number of individuals. MLG is the number of multilocus genotypes in the population. eMGL is the number of expected multilocus genotypes (i.e., genotypic richness) given the smallest sample size. λ is Simpson's index (genotypic diversity) after controlling for sample size. Hd is haplotype diversity. k indicates the number of unique haplotypes found in a population. Percent unique is the proportion of haplotypes that are unique to the population (k/N). Percent private haplotypes is the proportion of haplotypes found solely in that population.

TABLE 4 Within-Haplogroup Estimates of Diversity. Estimates of diversity across all haplotypes within one of the three most common haplogroups.

Population	R1a					R1b					I1				
	N	MLG	eMLG	λ	Hd	N	MLG	eMLG	λ	Hd	N	MLG	eMLG	λ	Hd
Faroes	58	19	6.78	0.89	0.89	35	23	8.29	0.94	0.94	29	11	6.07	0.86	0.86
Denmark	31	27	9.54	0.99	0.99	69	69	9.62	0.99	0.99	59	37	8.98	0.97	0.97
Ireland	2	2	2	1	1.00*	124	124	9.17	0.98	0.98	6	5	5	0.93	0.93
Iceland	35	19	8.03	0.95	0.95	73	73	8.95	0.97	0.97	52	24	6.48	0.83	0.83
Norway	11	10	9.18	0.98	0.98	15	15	9.14	0.98	0.98	13	9	7.38	0.92	0.92
Sweden	6	6	6	1	1	5	5	5	1	1	22	16	8.53	0.96	0.96
Total	143	83	9.18	0.98	1	321	321	9.76	0.99	1	181	102	9.21	0.98	0.96

N indicates the number of individuals. MLG is the number of multilocus genotypes in the population. eMGL is the number of expected multilocus genotypes (i.e., genotypic richness) given the smallest sample size. λ is Simpson's index (genotypic diversity) after controlling for sample size. Hd is haplotype diversity.

J2b (n = 4), J2a1b (n = 2), and I2a1 (n = 10). Of the rare haplogroups found among the Faroese, six were found in one or more assumed source populations. The Faroese haplogroups J1 and L were not found in any source population. We detected some inter-island variability between the Faroese sample populations, such that while haplogroup Q is shared among Tvorøyri, Klaksvík, and Tórshavn, haplogroups E1b1b, I2b, and I2a are found only on the northern island of Klaksvík, and haplogroups I2b1, L, and N are found only in the capital city of Tórshavn (Table 2). However, most of these haplogroups are represented by single individuals.
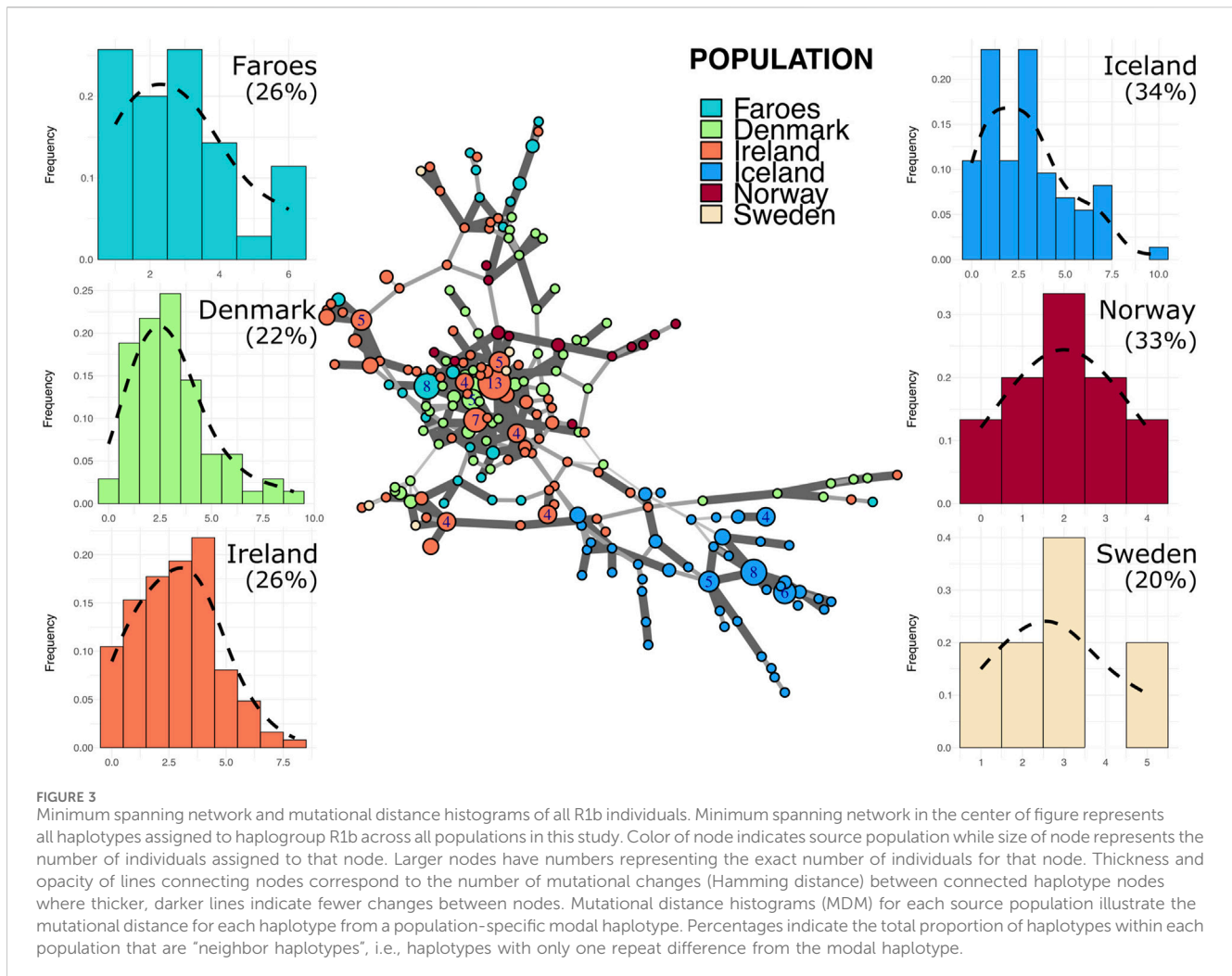
3.2 Faroese Y-chromosomal haplotypes

Like haplogroups, haplotypes tend to be more variable among mainland Scandinavian samples than in the Faroese and Icelandic samples. Across all haplogroups, the lowest levels of genotypic richness (eMLG) and diversity (λ) are found among the Faroese, Icelandic, and Irish populations but the levels of haplotype diversity (Hd) across all populations is relatively uniform with the lowest Hd found among the Faroes population and highest among the Danish population (Table 3). The Faroese population is the least diverse at the haplotype level, such that 49% of the Faroese haplotypes are unique, while private haplotypes constitute 30% of the sample. Of the total unique haplotypes found in the Faroese sample, 31% (21 haplotypes) are shared with one or more source populations. The highest number of haplotypes assigned to R1a were shared among the Faroese and Danish populations followed by the Faroes and Ireland, and the Faroes and Norway (Figure 2B). Among haplotypes assigned to R1b, the highest number were shared among the Faroes and Ireland, followed by the Faroes and Norway, and the Faroes and Denmark (Figure 2B). Finally, among haplotypes assigned to haplogroup I1, the Faroes shared the most haplotypes with Norway, followed by Sweden (Figure 2B). Considering all haplotypes independent of haplogroup identity, the sample from the Faroes clusters closely with Norway, Denmark, and to a lesser extent, Sweden using correspondence analysis (Figure 2C).

3.3 Within−haplogroup haplotype diversity

Unlike estimates of haplotype diversity that do not consider haplogroup membership, within-haplogroup haplotype diversities vary substantially depending on the haplogroup in question. For example, while the Faroese and Icelandic populations have the

FIGURE 3
Minimum spanning network and mutational distance histograms of all R1b individuals. Minimum spanning network in the center of figure represents all haplotypes assigned to haplogroup R1b across all populations in this study. Color of node indicates source population while size of node represents the number of individuals assigned to that node. Larger nodes have numbers representing the exact number of individuals for that node. Thickness and opacity of lines connecting nodes correspond to the number of mutational changes (Hamming distance) between connected haplotype nodes where thicker, darker lines indicate fewer changes between nodes. Mutational distance histograms (MDM) for each source population illustrate the mutational distance for each haplotype from a population-specific modal haplotype. Percentages indicate the total proportion of haplotypes within each population that are "neighbor haplotypes", i.e., haplotypes with only one repeat difference from the modal haplotype.
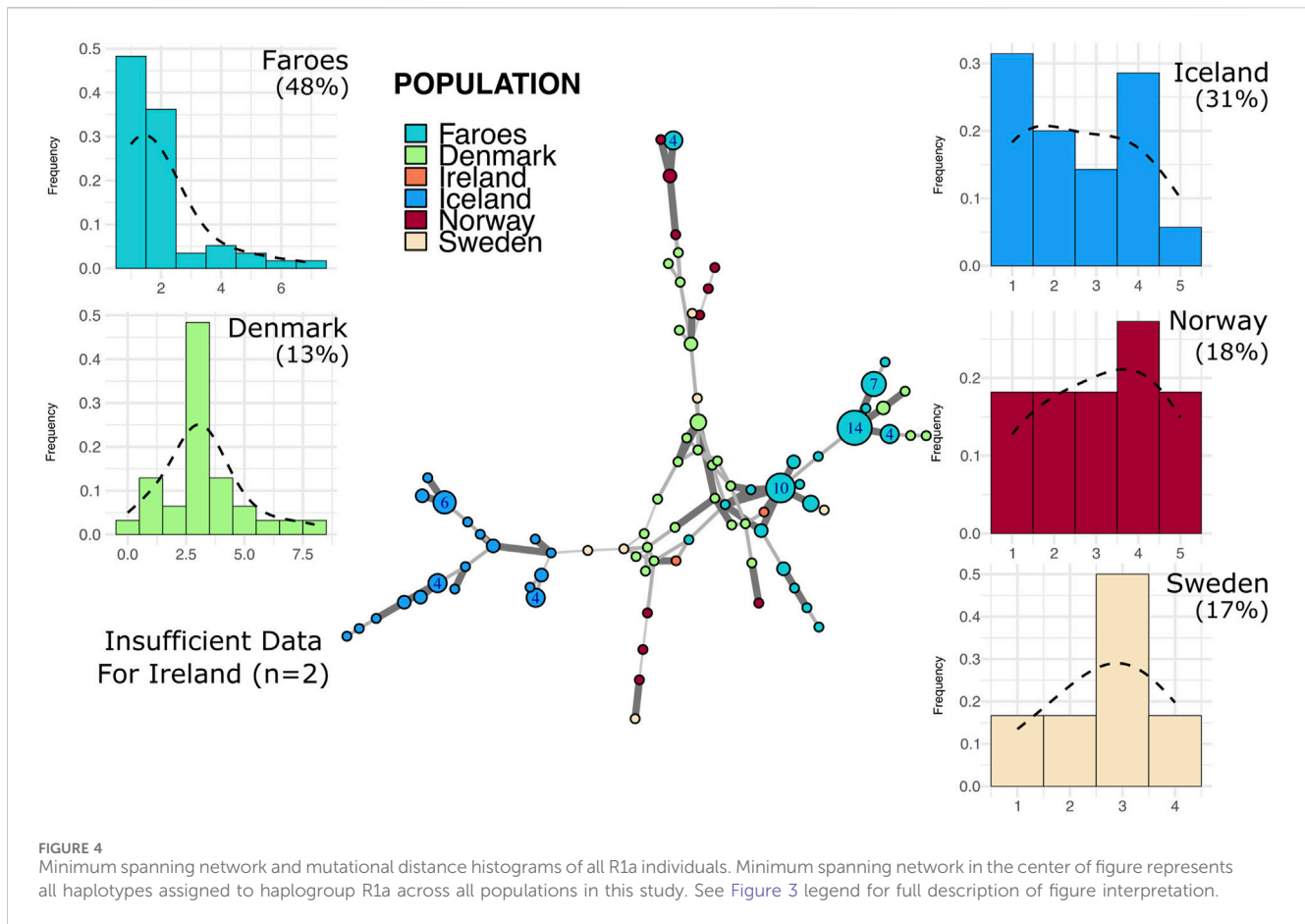
lowest within-haplogroup haplotype diversity for R1a and I1 haplogroups, within-haplogroup haplotype diversity metrics among R1b individuals are relatively high across all populations (Table 4). Notably, however, the lowest haplotype diversity is found among the Faroese and Icelandic haplotypes assigned to R1b. The low diversity observed among the Faroese and Icelandic populations is also reflected in the networks and histograms of mutational distances presented below.

3.4 Identification of founder effect in haplotype networks

In the R1b haplogroup network (Figure 3), several Faroese haplotypes stem from one larger common node, with other haplotypes dispersed across the network. The R1b network is dominated by the Irish sample with no population exclusively marking a cluster except for Iceland. Within the R1b haplogroup, Faroese haplotypes are shared with those found in Ireland, Norway, and Denmark (Figure 2B) but never with Sweden or Iceland. Danish R1b haplotypes have the highest within population distance of all source populations with a mean of 4.0 (±1.55). Mean genetic distance is also high among Swedish haplotypes (4.4, ±1.65) but

this is largely due to the small sample size of Swedish haplotypes assigned to this haplogroup (n = 5). For all other populations, the mean genetic distance is comparable: Faroes (3.54 ± 1.71), Ireland (3.76 ± 1.65), Iceland (3.79 ± 1.82), and Norway (3.04 ± 1.41). However, Faroese haplotype diversity within the R1b haplogroup (Table 3) is higher than that found in R1a and I1, implying that R1b haplotypes may have been more diverse in the original founding population. Alternatively, post-founding gene flow could have contributed to this higher degree of diversity (i.e., post-founding admixture with mainland Scandinavian populations). Among all major haplogroups, the degree of genetic differentiation between populations, while significant (p = 0.01), is lowest among R1b haplogroups ($\Phi PT = 0.29$) and the majority of observed genetic variation is within populations ($\Phi PT = 0.71$).

The R1a haplogroup network (Figure 4) is sparser as compared to the R1b haplogroup, primarily due to the lack of representation of our Irish population within this group. In the R1a haplogroup network, the Faroese population is characterized by several large nodes that primarily cluster with Danish haplotypes with a smaller and more distinct node that clusters with Norwegian R1a haplotypes. Like with haplogroup R1b, Faroese R1a haplotypes are shared with Denmark, Ireland, and Norway, but not with Sweden or Iceland. The mean genetic distance among all Faroese

FIGURE 4
Minimum spanning network and mutational distance histograms of all R1a individuals. Minimum spanning network in the center of figure represents all haplotypes assigned to haplogroup R1a across all populations in this study. See Figure 3 legend for full description of figure interpretation.

R1a haplotypes was substantially lower than any other source population (2.55 ± 1.62). Swedish R1a haplotypes had the highest mean genetic distance (4.6 ± 0.99) followed by Norway (4.35 ± 1.65), Denmark (4.04 ± 1.51), and Iceland (3.08 ± 1.35). Too few individuals were assigned to this haplogroup in Ireland to perform mean distance analysis (n = 2). Genetic differentiation among individuals is higher among R1a haplotypes (ΦPT = 0.34, p = 0.01) than R1b haplotypes but most of the genetic variation is similarly found within populations (ΦPT = 0.66).

Most Faroese haplotypes in the I1 network (Figure 5) are closely clustered with two larger nodes and several smaller nodes that all fall central to other source populations with the exception, again, of Iceland. Additionally, Faroese I1 haplotypes are shared with all source populations with the exception of Iceland (Figure 2B). In addition to having the lowest haplotype diversity within the I1 haplogroup (0.86, Table 3), the Faroese I1 haplotypes have the lowest mean genetic distance among all populations (1.83 ± 1.36) followed by Iceland (2.18 ± 1.58), Norway (2.79 ± 1.51), Ireland (3.13 ± 1.41), Sweden (3.23 ± 1.63), and Denmark (3.42 ± 1.7). We detected the highest degree of population differentiation among I1 haplotypes wherein 44% (ΦPT = 0.44, p = 0.01) of all genetic variation was found between populations and only 56% (ΦPT = 0.56) within populations.
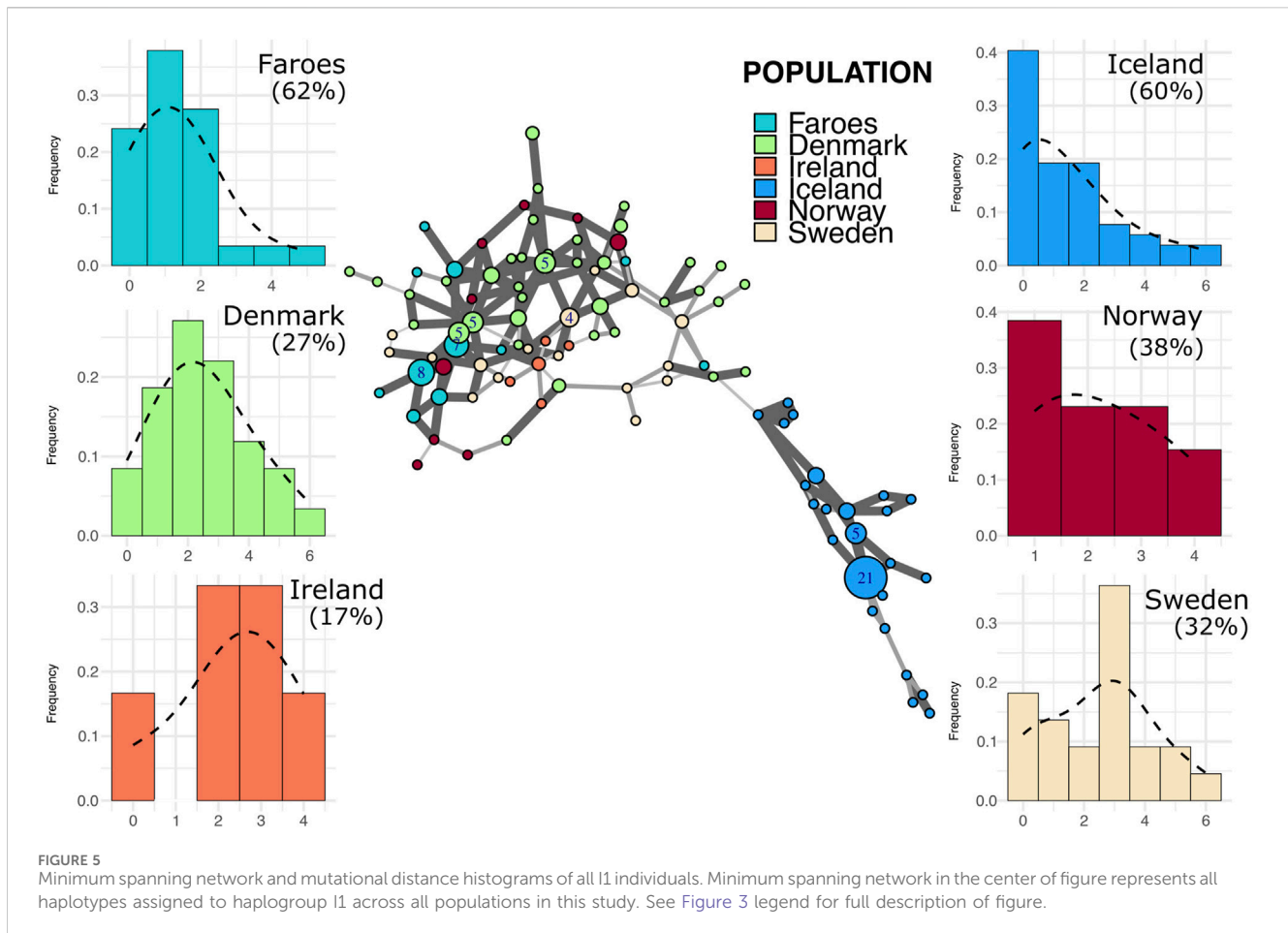
To summarize, the distribution of Faroese haplotypes across minimum spanning networks within haplogroups do not match the stereotypical patterns for recently founded populations; specifically, the presence of clusters of singleton haplotypes surrounding large

central nodes in star-like formations. However, we do find differences in population differentiation across major haplogroups with different degrees of observed genetic variation within and between groups, supporting the idea that within haplogroup diversity estimates may have finer resolution for documenting population events such as founder effect.

3.5 Identification of founder effect in the distribution of mutational distances

Histograms representing stepwise mutational distances from within-haplogroup modal haplotypes for each population are found in Figures 3–5. The results of this analysis reveals two basic patterns among all populations, mainly smooth distributions with a single mode and multi-modal distributions.

Putative source populations (i.e., Denmark, Ireland, Norway, and Sweden), generally have MDM distributions characterized by a single mode, but the mode has shifted further away from zero with a lower overall proportion of neighbor haplotypes, indicating the kind of accumulated variability expected in older populations. In contrast, the Faroese and Icelandic samples tend to have modes nearer to zero with a higher proportion of neighbor haplotypes and are skewed to the right. This pattern, however, is haplogroup specific which may reflect differences in the diversity of the original founding group, admixture post founding, or other aspects of population structure. For example, the distribution of haplotypes

FIGURE 5
Minimum spanning network and mutational distance histograms of all I1 individuals. Minimum spanning network in the center of figure represents all haplotypes assigned to haplogroup I1 across all populations in this study. See Figure 3 legend for full description of figure.

from the modal haplotype within haplogroup I1 is highly skewed towards the right among the Faroese ($\gamma1 = 1.19$) and Icelandic populations ($\gamma1 = 1.13$) with 60% or more haplotypes representing neighbors to the mode (Figure 5). Similarly, Faroese haplotypes are strongly skewed to the right within the R1a haplogroup ($\gamma1 = 2.02$) with 48% of all haplotypes representing neighboring haplotypes. Interestingly, while a relatively high proportion of Icelandic haplotypes within the R1a haplogroup are neighbors to the mode (31%), the distribution of Icelandic R1a haplotypes is multimodal ($\gamma1 = 0.17$) which reflects underlying population structure also seen in our minimum spanning networks. This difference in population structure is again clearly seen among MDM histograms representing R1b haplotypes where there is a clear multimodal distribution that is skewed towards the right among both the Faroese ($\gamma1 = 0.63$) and Icelandic ($\gamma1 = 0.78$) haplotypes but a smooth accumulation of mutational distance from the mode among Danish ($\gamma1 = 1$), Irish ($\gamma1 = 0.28$), Norwegian ($\gamma1 = 0$), and Swedish populations ($\gamma1 = 0.37$).

To test whether these multimodal distributions found among the Faroes and Icelandic R1b and R1a haplotypes are the result of true within-haplogroup diversity or haplogroup classification error, we performed an additional "strict" haplogroup MDM histogram analysis where to be included a haplotype must have had a probability score above 90% and a fitness score of >50 using the Athey haplogroup prediction program. Notably, however, the proportion of neighbor haplotypes among all populations are

similar despite a clear shift from the modal haplotype among our North Atlantic Island populations. Interestingly, while the multimodal distribution previously observed among the Faroese and Icelandic R1a haplotypes diminished significantly using these strict parameters, the multimodal distribution found among the Faroese R1b haplogroup remained relatively unchanged (Supplementary Figure S2).

Despite variation among different haplogroups, these results follow the expectation that source populations have broader and smoother distributions, notwithstanding sample size and the effect of latent population structure. In contrast, the MDM histograms from the Faroese sample have modes nearer zero and are skewed to the right.

4 Discussion

The Vikings played a preeminent role in the peopling of the North Atlantic, and one might expect populations that were founded by the Vikings to be genetically similar and homogeneous. Previous investigation of Y–chromosomal haplotype and haplogroup frequencies in the Faroe Islands found a high degree of similarity between the Faroe Islands, Norway, Sweden, and Iceland, indicating a Scandinavian ancestry among the male settlers of the Faroe Islands (Jorgensen et al., 2004). Faroese mitochondrial lineages, however, indicate an excess of British Isles ancestry among the female settlers

of the Faroe Islands (Als et al., 2006). In both studies the Faroese population was found to be highly homogeneous, reflecting an expected small founder population size and strong genetic drift. In the case of the Y–chromosome, there were relatively few haplotypes per sample, relatively low gene diversity, and a correspondingly low effective population size (Jorgensen et al., 2004). In the current study we argue that the application of a moderately extended haplotype in conjunction with consideration of estimates of diversity in the context of network analysis can more clearly elucidate founder effect in the Faroese population. Specifically, we argue that the opportunity to detect loss of diversity associated with founder effect will be affected by the level of phylogenetic resolution considered. The significance of hierarchical analysis for this study is that estimates of haplotype diversity may be inflated by underlying haplogroup diversity and the degree to which those haplogroups have diverged in conjunction with their position on the Y–chromosomal gene tree. Thus, an improved assessment of hierarchical levels of diversity enhances our understanding of the genetic history of the Faroe Islands and the consequences of the founder event.

4.1 Private and unique haplotypes

The most constrained genetic diversity is consistently found in the Faroese samples. The sample from the Faroe Islands has the lowest percentages of unique haplotypes (i.e., the total number of unique haplotypes within a population) as well as private haplotypes (i.e., haplotypes found in that population while completely absent from all other populations). This is consistent with a comparatively recent small-scale founding event. Specifically, unique haplotypes are frequently found as a result of demographic expansion (Slatkin and Hudson, 1991). Similarly, the frequency of private haplotypes found in a given population can be influenced by the degree of sampling saturation, divergence from other populations, and genetic drift (Helgason et al., 2000). The presence of private haplotypes is also a function of genetic resolution, such that the more microsatellites scored, the more likely a given haplotype will be private within a population. Additionally, differences of private haplotypes may also be partly the consequence of sampling strategy. Population structure can be found in large, geographically dispersed source populations, and is best captured with broad geographic sampling [cf. Dupuy et al. (2006) for the example of Norway]. In the present study, the broad geographic sampling scheme for Sweden may explain minimal sharing of haplotypes (6%), while Jorgensen et al. (2004) reported 34% of Faroese Y–chromosomes were assigned to Sweden.

4.2 Parsing haplotype diversity

Measurements of diversity will be affected by whether haplotype variability is considered without respect to haplogroup. The significance of this is that haplotype diversity estimates will be inflated depending on the degree of haplogroup diversity found in the population: Disregarding haplogroup identity will obfuscate a signal of founder effect. Each new haplogroup represents a different evolutionary history, and it is inappropriate to look at microsatellite

diversity by combining data from across many different haplogroup lineages. In the present study, haplotype diversity estimates from island populations are comparable to source populations when haplotypes are considered without respect to haplogroup, yet within–haplogroup haplotype diversity estimates are substantially lower (Table 3). This line of reasoning likely extends to other summary statistics.

4.3 Faroese haplotypes are distributed on haplogroup networks

It is a general expectation that star–like clusters of singleton haplotypes crowded around large primary nodes in haplotype networks are concordant with founder events and subsequent expansion. This has been demonstrated in Y–chromosomal studies of Ashkenazi Jews (Nebel et al., 2005), the population living on the Indonesian island of Nias (van Oven et al., 2011), and in the Irish (Moore et al., 2006). In addition to some primary nodes, several Faroese haplotypes are found widely dispersed on the networks. The position and distribution of Faroese haplotypes across networks depends on certain variables: the effective population size of the initial founders in conjunction with the impact of genetic drift since the founding event; the relative contribution of more or less divergent haplotypes to the founding population; and the degree of reproductive isolation since the time of founding. If the original effective size was small, we would expect to see a few larger nodes with closely related haplotypes clustered nearby. If paternal contributions are geographically and phylogenetically diverse, however, Faroese haplotypes may show up scattered across each network. In the present study, we find generally that singleton Faroese haplotypes are widely distributed across each haplogroup network, the majority of which are found either in large clusters of primary haplotype nodes and shared with one or more source populations, or close one–and two–step neighbor haplotype nodes. The distributions of Faroese haplotypes across the haplogroup networks demonstrate that the Faroese colonizers likely originated from multiple source populations. These results are consistent with recent findings that the Viking world was highly genetically diverse and that Scandinavia during the Viking age experienced significant genetic influx from divergent European regions (Krzewińska et al., 2015; Margaryan et al., 2020). This genomic history is likely reflected in contemporary populations that were colonized during the Viking diaspora, including the Faroes.

4.4 Icelandic haplotypes are isolated on networks

Although within–haplogroup haplotype diversity estimates in samples from the Faroe Islands and Iceland are comparatively lower than source populations, the degree of Icelandic haplotype isolation in each network is distinctive. For each of the haplogroup networks, most Icelandic haplotypes are tightly clustered to the exclusion of haplotypes from other populations; these isolated clusters of haplotypes are not shared with any source populations or the Faroe Islands. Both

the Faroese and Icelandic founders are thought to have been mostly Vikings who originated primarily from Norway, Sweden, or Denmark, and the distribution of Faroese haplotypes on the networks demonstrates this relationship. That the isolated clusters of Icelandic haplotypes are not shared may be explained by incomplete sampling of source populations, namely that the region(s) where Icelandic founding haplotypes originated was not sampled. A larger, more diverse set of source population samples with greater regional differentiation may clarify the respective origins of the Faroese and Icelandic founder groups. It is also plausible that the Icelandic founder population originated from long–established Viking colonies in the British Isles (e.g., Lall et al., 2021), while the Faroese are the descendants of a more direct Scandinavian settlement process. Regarding the initial Icelandic settlers, such a scenario was proposed by Helgason et al. (2000). Thus, successive and serial bottlenecking of genetic variability of the Icelanders coming by way of the British Isles may explain both their position and stark clustering in the haplogroup networks.

4.5 Mutation distance from the modal haplotype histograms

Faroese and Icelandic R1a and I1 haplogroup MDM histograms are typical of our expectations for a signal of founder effect, that is, highly leptokurtic distributions which are skewed to the right. As the modal haplotype is constructed by calculating the mode for each locus and combining them, it is expected that, in a population recently founded, most individuals within the haplogroup will share the modal haplotype. These MDM histograms reveal a recent, small–scale founder event in that most haplotypes are typically fewer than two mutational steps away from the calculated mode. The Faroese pattern is distinct from that of the source populations where the central tendency of a given source population is often at two or more mutational steps from the calculated mode.

These results contrast somewhat with those of haplogroup R1b. Bearing in mind that R1b is an older haplogroup than both R1a and I1 (Hallast et al., 2015), greater haplotype diversity is expected. In the case of the Faroe Islands, the most notable feature is a strong multimodal distribution with a relatively high number of haplotypes that are six mutational steps from the calculated mode. Similarly, Iceland is multimodal with a relatively high proportion of haplotypes that are highly divergent from the mode, a pattern similar to that seen in the Danish and Irish R1b haplotypes. It is possible that this pattern is the result of convergence during founding of two distinct branches contributing divergent haplotypes to the original R1b founders of Iceland and the Faroes (cf. Helgason et al., 2000). Alternatively, these patterns among contemporary Faroese may indicate more recent admixture from Denmark. The Faroe Islands are today a self-governing part of the Danish Realm. For centuries, however, there has been a close relationship between Faroe Islands and Denmark, mainly political and administrative, but also on family and personal levels. It is likely that this shared political and demographic history led to gene flow between the two countries.

Although most of the Faroese population belongs to one of three highly frequent haplogroups, there are some rare haplogroups present in the population that are infrequent even among Europeans as a whole, and local frequencies of rare haplogroups vary between Faroese locales (Table 2). In addition, specific haplogroups found in this sample are likely indicative of migration after the initial founding of the Faroe Islands. This assertion is based on the expectation that haplogroups present in the initial founders should be at higher frequency. The only rare haplogroup of note is J1, which constitutes 10% of the Klaksvík Y chromosomes. Haplogroup J is thought to have originated in the Middle East and is associated with the origins and spread of early farming groups. Unlike the European J–M172 sub–haplogroup, however, J1 is most closely associated with Middle Eastern, northern African, and Ethiopian populations (Karafet et al., 2008). Haplogroup J1 is not found in any source population investigated in this study and is therefore highly unlikely to have originated in the founding group. Perhaps most intriguing is historical evidence that may account for this haplogroup's presence in the Faroe Islands, specifically an attack by the Barbary corsairs in the 1600s (West and Jákupsson, 1980). Stories about Turkish pirates who came to the southernmost islands, Suðuroy, are well known among Faroese peoples. Modern populations living in regions where Barbary corsair ports once existed still possess high frequencies of the J1 haplogroup (Arredi et al., 2004). The J haplogroup is also found at higher than expected frequencies in other Atlantic island populations with complex colonization histories. For example, an investigation into Y–chromosomal diversity in the Azores revealed that J haplogroup frequencies were twice those in mainland Portugal (Pacheco et al., 2005). It is also possible that the presence of haplogroup L in the Faroes is similarly a more recent introduction. However, as the haplogroup L is found in only a single individual, and is the basal level of the haplogroup, interpretation of the presence of this group in the Faroes is difficult. Regardless of the source, haplogroups that are rare within the overall Faroe Islands sample inflate haplotype diversity when estimates are calculated without regard to haplogroup identity. This follows from the observation that haplotypes of regionally uncommon haplogroups will be more divergent from more common haplogroups and thus have a higher probability of being both private and unique.

4.6 Faroese-Icelandic comparison

Although little is actually known about the settlement process of the Faroe Islands, it has largely been assumed that it was similar to that of Iceland (Arge et al., 2005), and historical data document the possibility that the Faroese and Icelanders share common founders. Previous genetic studies using lower resolution Y–chromosomal data have found close connections between the Faroese and Icelanders (Jorgensen et al., 2004). It has been estimated that Y–chromosomal lines of descent can be reliably discerned over 1,950 generations (or 49,000 years) (de Knijff, 2000). Therefore, if Iceland and the Faroe Islands do share a common paternal source—or if there has been post–founder contact between the two populations—the data should reflect this at the haplotype level within major haplogroups. The data presented here, however, provide little evidence to support the

hypothesis that the Faroese and Icelanders originated from the same proximate source populations or have undergone subsequent episodes of admixture.

This lack of obvious concordance of founders between the Faroe Islands and Iceland in the current study is unexpected. We posit that the discrepancy between the current and previous studies comparing the Faroes and Iceland is due to the number of microsatellite loci used in the original study (five) and that of the current study (nine of the 12 loci presented here). Divergence can be more easily detected if more Y microsatellite loci are available for comparison, and these loci are relatively more mutable. The high level of clustering and isolation on the networks of the Icelandic samples is evocative of a latent SNP present in the sample. Because this study is confined to a particular level of haplogroup resolution, the representation of latent SNPs in the data can only be inferred and not implied directly. If the haplogroup defined by the latent SNP is sufficiently divergent from other haplogroup members and there are enough representative individuals within the data, clustering of a subgroup may be apparent within the larger haplogroup itself. This method has been used to distinguish sub–haplogroups in Northern Ireland (Moore et al., 2006) and particularly successful male lineages across Asia (Zerjal et al., 2003).

If Iceland and the Faroe Islands are indeed genetically divergent, as found in the present study, two scenarios may explain their differentiation and Iceland's relative isolation on the networks: (1) the populations were settled by different regional groups; or, (2) another historically contingent process of evolution between the two groups has resulted in the divergence. In the case of the former, remote, geographically divergent Norwegian sources may have generated the multimodal distribution evident in R1b (Dupuy et al., 2006). In the case of the latter, as mentioned previously, and proposed by Helgason et al. (2000), multiple bottlenecks of genetic variation of Vikings settling in Ireland and those same Vikings colonizing Iceland may also lead to isolation on networks. These scenarios are not mutually exclusive, and our data lack sufficient resolution to distinguish between them. Both scenarios are consistent with reports of paucity of Icelandic genetic variation (as a function of genetic drift) (Helgason et al., 2003) and evidence for admixture incorporating divergent lineages from Scandinavia and Celtic populations giving rise to deeper heterogeneity even in the context of founder effect (Helgason et al., 2000; Arnason, 2003).

Regardless, our results suggest that the Icelandic and Faroese populations had distinguishably different founding fathers. Our results are consistent with the scenario that the Faroese male population was founded by a more diverse group from divergent Scandinavian populations than their Icelandic neighbors. Furthermore, we conclusively demonstrate that there is no evidence for post–founder admixture between the Faroese and Icelandic Y–chromosomal gene pools.

Data availability statement

The datasets presented in this study can be found in online repositories. The names of the repository/repositories and accession number(s) can be found in the article/ Supplementary Material.

Ethics statement

Ethical approval for the Faroese samples was waived by the Scientific Ethics Commission of the Faroe Islands, as the study was deemed anthropological in nature. At the time of sample collection, the Commission's jurisdiction covered only biomedical research, excluding anthropological studies. The Norwegian and Swedish human samples, collected in 1999, were obtained under the ethical guidelines of the University of Arizona, which did not require Institutional Review Board (IRB) approval for anonymous, population-based samples. All samples were fully anonymized before analysis, with no identifying information gathered. Current ethical guidelines for human subject research were considered, and re-identification of the samples is not possible. The research was conducted in accordance with local laws and institutional requirements. Written informed consent was provided by the participants or their legal guardians/next of kin.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2024.1462736/full#supplementary-material

References

Als, T. D., Jorgensen, T. H., Børglum, A. D., Petersen, P. A., Mors, O., and Wang, A. G. (2006). Highly discrepant proportions of female and male Scandinavian and British Isles ancestry within the isolated population of the Faroe Islands. *Eur. J. Hum. Genet.* 14, 497–504. doi:10.1038/sj.ejhg.5201578

Arge, S. V. (1991). The landnám in the Faroes. *Artic. Anthropol.* 28 (2), 101–120.

Arge, S. V., Sveinbjarnardóttir, G., Edwards, K. J., and Buckland, P. C. (2005). Viking and medieval settlement in the Faroes: people, place and environment. *Hum. Ecol.* 33, 597–620. doi:10.1007/s10745-005-4745-1

Arnason, E. (2003). Genetic heterogeneity of Icelanders. *Ann. Hum. Genet.* 67 (Pt 1), 5–16. doi:10.1046/j.1469-1809.2003.00003.x

Arredi, B., Poloni, E. S., Paracchini, S., Zerjal, T., Fathallah, D. M., Makrelouf, M., et al. (2004). A predominantly neolithic origin for Y-chromosomal DNA variation in North africa. *Am. J. Hum. Genet.* 75, 338–345. doi:10.1086/423147

Athey, T. W. (2006). Haplogroup prediction from Y-STR values using a Bayesian-allele-frequency approach. *J. Genet. Geneal.* 2 (2), 34–39.

Ballard, D. J., Phillips, C., Thacker, C. R., and Court, D. S. (2006). Y chromosome STR haplotype data for an Irish population. *Forensic Sci. Int.* 161, 64–68. doi:10.1016/j.forsciint.2005.08.010

Church, M. J., Arge, S. V., Edwards, K. J., Ascough, P. L., Bond, J. M., Cook, G. T., et al. (2013). The Vikings were not the first colonizers of the Faroe Islands. *Quat. Sci. Rev.* 77, 228–232. doi:10.1016/j.quascirev.2013.06.011

Curtin, L., D'Andrea, W. J., Balascio, N. L., Shirazi, S., Shapiro, B., De Wet, G. A., et al. (2021). Sedimentary DNA and molecular evidence for early human occupation of the Faroe Islands. *Commun. Earth Environ.* 2, 253. doi:10.1038/s43247-021-00318-0

de Knijff, P. (2000). Messages through bottlenecks: on the combined use of slow and fast evolving polymorphic markers on the human Y chromosome. *Am. J. Hum. Genet.* 7.

Di Rienzo, A., and Wilson, A. C. (1991). Branching pattern in the evolutionary tree for human mitochondrial DNA. *Proc. Natl. Acad. Sci.* 88, 1597–1601. doi:10.1073/pnas.88.5.1597

Dray, S., and Dufour, A.-B. (2007). The ade4 package: implementing the duality diagram for ecologists. *J. Stat. Softw.* 22, 1–20. doi:10.18637/jss.v022.i04

Dupuy, B. M., Stenersen, M., Lu, T. T., and Olaisen, B. (2006). Geographical heterogeneity of Y-chromosomal lineages in Norway. *Forensic Sci. Int.* 164, 10–19. doi:10.1016/j.forsciint.2005.11.009

Gislason, H. (2023). SNP heterozygosity, relatedness and inbreeding of whole genomes from the isolated population of the Faroe Islands. *BMC Genomics* 24, 707. doi:10.1186/s12864-023-09763-x

Goodacre, S., Helgason, A., Nicholson, J., Southam, L., Ferguson, L., Hickey, E., et al. (2005). Genetic evidence for a family-based scandinavian settlement of shetland and orkney during the viking periods. *Heredity* 95, 129–135. doi:10.1038/sj.hdy.6800661

Hallast, P., Batini, C., Zadik, D., Maisano Delser, P., Wetton, J. H., Arroyo-Pardo, E., et al. (2015). The Y-chromosome tree bursts into leaf: 13,000 high-confidence SNPs covering the majority of known clades. *Mol. Biol. Evol.* 32, 661–673. doi:10.1093/molbev/msu327

Hallenberg, C., Nielsen, K., Simonsen, B., Sanchez, J., and Morling, N. (2005). Y-chromosome STR haplotypes in Danes. *Forensic Sci. Int.* 155, 205–210. doi:10.1016/j.forsciint.2004.12.019

Helgason, A., Nicholson, G., Stefansson, K., and Donnelly, P. (2003). A reassessment of genetic diversity in Icelanders: strong evidence from multiple loci for relative homogeneity caused by genetic drift. *Ann. Hum. Genet.* 67, 281–297. doi:10.1046/j.1469-1809.2003.00046.x

Helgason, A., Sigurðardóttir, S., Nicholson, J., Sykes, B., Hill, E. W., Bradley, D. G., et al. (2000). Estimating scandinavian and gaelic ancestry in the male settlers of Iceland. *Am. J. Hum. Genet.* 67, 697–717. doi:10.1086/303046

Johnston, G. (1975). *The Faroe islanders' saga*. Ottawa: Oberon Press.

Jombart, T. (2008). adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics* 24, 1403–1405. doi:10.1093/bioinformatics/btn129

Jones, G. (1984). *A history of the Vikings*. New York: Oxford University Press.

Jorgensen, T. H., Buttenschon, H. N., Wang, A. G., Als, T. D., Borglum, A. D., and Ewald, H. (2004). The origin of the isolated population of the Faroe Islands investigated using Y chromosomal markers. *Hum. Genet.* 115, 19–28. doi:10.1007/s00439-004-1117-7

Kamvar, Z. N., Tabima, J. F., and Grünwald, N. J. (2014). Poppr: an R package for genetic analysis of populations with clonal, partially clonal, and/or sexual reproduction. *PeerJ* 2, e281. doi:10.7717/peerj.281

Karafet, T. M., Mendez, F. L., Meilerman, M. B., Underhill, P. A., Zegura, S. L., and Hammer, M. F. (2008). New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. *Genome Res.* 18, 830–838. doi:10.1101/gr.7172008

Kayser, M., Caglià, A., Corach, D., Fretwell, N., Gehrig, C., Graziosi, G., et al. (1997). Evaluation of Y-chromosomal STRs: a multicenter study. *Int. J. Leg. Med.* 110, 125–133. doi:10.1007/s004140050051

Kimura, M., and Ohta, T. (1978). Stepwise mutation model and distribution of allelic frequencies in a finite population. *Proc. Natl. Acad. Sci.* 75, 2868–2872. doi:10.1073/pnas.75.6.2868

Komsta, L., and Novomestky, F. (2022). Moments: moments, cumulants, skewness, kurtosis and related tests. Available at: https://cran.r-project.org/web/packages/moments/index.html (Accessed May 20, 2024).

Krzewińska, M., Bjørnstad, G., Skoglund, P., Olason, P. I., Bill, J., Götherström, A., et al. (2015). Mitochondrial DNA variation in the Viking age population of Norway. *Philos. Trans. R. Soc. B Biol. Sci.* 370, 20130384. doi:10.1098/rstb.2013.0384

Lall, G. M., Larmuseau, M. H. D., Wetton, J. H., Batini, C., Hallast, P., Huszar, T. I., et al. (2021). Subdividing Y-chromosome haplogroup R1a1 reveals Norse Viking dispersal lineages in Britain. *Eur. J. Hum. Genet.* 29, 512–523. doi:10.1038/s41431-020-00747-z

Magnusson, M. (2003). The Vikings. *Stroud Tempus*.

Margaryan, A., Lawson, D. J., Sikora, M., Racimo, F., Rasmussen, S., Moltke, I., et al. (2020). Population genomics of the Viking world. *Nature* 585, 390–396. doi:10.1038/s41586-020-2688-8

Moore, L. T., McEvoy, B., Cape, E., Simms, K., and Bradley, D. G. (2006). A Y-chromosome signature of hegemony in gaelic Ireland. *Am. J. Hum. Genet.* 78, 334–338. doi:10.1086/500055

Nebel, A., Filon, D., Faerman, M., Soodyall, H., and Oppenheim, A. (2005). Y chromosome evidence for a founder effect in Ashkenazi Jews. *Eur. J. Hum. Genet.* 13, 388–391. doi:10.1038/sj.ejhg.5201319

Oksanen, J., Blanchet, F. G., Friendly, M., Kiindt, R., Legendre, P., McGlinn, D., et al. (2019). Vegan: community ecology package. Available at: https://CRAN.R-project.org/package=vegan.

Pacheco, P. R., Branco, C. C., Cabral, R., Costa, S., Araújo, A. L., Peixoto, B. R., et al. (2005). The Y-chromosomal heritage of the Azores islands population: Y-chromosome in the Azores islands (Portugal). *Ann. Hum. Genet.* 69, 145–156. doi:10.1046/j.1529-8817.2004.00147.x

Paradis, E. (2010). pegas: an R package for population genetics with an integrated–modular approach. *Bioinformatics* 26, 419–420. doi:10.1093/bioinformatics/btp696

Powell, F. Y. (1896). *The tale of thrond of gate: commonly called faereyinga saga*. London: Ballantyne Press.

Quintana-Murci, L., Bigham, A., Rouba, H., Barakat, A., McElreavey, K., and Hammer, M. (2004). Y-chromosomal STR haplotypes in Berber and Arabic-speaking populations from Morocco. *Forensic Sci. Int.* 140, 113–115. doi:10.1016/j.forsciint.2003.11.006

Redd, A. J., Agellon, A. B., Kearney, V. A., Contreras, V. A., Karafet, T., Park, H., et al. (2002). Forensic value of 14 novel STRs on the human Y chromosome. *Forensic Sci. Int.* 130, 97–111. doi:10.1016/S0379-0738(02)00347-X

Rogers, A., and Harpending, H. (1992). Population growth makes waves in the distribution of pairwise genetic differences. *Mol. Biol. Evol.* 9, 552–569. doi:10.1093/oxfordjournals.molbev.a040727

Simpson, E. H. (1949). Measurement of diversity. *Nature* 163, 688. doi:10.1038/163688a0

Slatkin, M., and Hudson, R. R. (1991). Pairwise comparisons of mitochondrial DNA sequences in stable and exponentially growing populations. *Genetics* 129, 555–562. doi:10.1093/genetics/129.2.555

Strøm, H. P. (2017). "The demographic development of the Faroe Islands during the last 300 years [Demografiska menningin í Føroyum síðstu 300 árini]," in *Statistics Faroe Islands [hagstova førpya], (argir)*, 43.

The Y Chromosome Consortium (2002). A nomenclature system for the tree of human Y-chromosomal binary haplogroups. *Genome Res.* 12, 339–348. doi:10.1101/gr.217602

van Oven, M., Hammerle, J. M., van Schoor, M., Kushnick, G., Pennekamp, P., Zega, I., et al. (2011). Unexpected island effects at an extreme: reduced Y chromosome and mitochondrial DNA diversity in Nias. *Mol. Biol. Evol.* 28, 1349–1361. doi:10.1093/molbev/msq300

West, J., and Jákupsson, B. (1980). *Faroese folk-tales and legends*. Lerwick, Shetland: Shetland Publishing Company.

Willuweit, S., and Roewer, L. (2015). The new Y Chromosome haplotype reference Database. *Forensic Sci. Int. Genet.* 15, 43–48. doi:10.1016/j.fsigen.2014.11.024

Wilson, J. F., Weiss, D. A., Richards, M., Thomas, M. G., Bradman, N., and Goldstein, D. B. (2001). Genetic evidence for different male and female roles during cultural transitions in the British Isles. *Proc. Natl. Acad. Sci.* 98, 5078–5083. doi:10.1073/pnas.071036898

Xue, Y., Wang, Q., Long, Q., Ng, B. L., Swerdlow, H., Burton, J., et al. (2009). Human Y chromosome base-substitution mutation rate measured by direct sequencing in a deep-rooting pedigree. *Curr. Biol.* 19, 1453–1457. doi:10.1016/j.cub.2009.07.032

Zerjal, T., Xue, Y., Bertorelle, G., Wells, R. S., Bao, W., Zhu, S., et al. (2003). The genetic legacy of the mongols. *Am. J. Hum. Genet.* 72, 717–721. doi:10.1086/367774

Zhivotovsky, L. A., Underhill, P. A., Cinnioğlu, C., Kayser, M., Morar, B., Kivisild, T., et al. (2004). The effective mutation rate at Y chromosome Short Tandem repeats, with application to human population-divergence time. *Am. J. Hum. Genet.* 74, 50–61. doi:10.1086/380911